

The Pathway Editor: A tool for managing complex biological networks

A. Sorokin
K. Paliy
A. Selkov
O. V. Demin
S. Dronov
P. Ghazal
I. Goryanin

Biological networks are systems of biochemical processes inside a cell that involve cellular constituents such as DNA, RNA, proteins, and various small molecules. Pathway maps are often used to represent the structure of such networks with associated biological information. Several pathway editors exist, and they vary according to specific domains of knowledge. This paper presents a review of existing pathway editors, along with an introduction to the Edinburgh Pathway Editor (EPE). EPE was designed for the annotation, visualization, and presentation of a wide variety of biological networks that include metabolic, genetic, and signal transduction pathways. EPE is based on a metadata-driven architecture. The editor supports the presentation and annotation of maps, in addition to the storage and retrieval of reaction kinetics information in relational databases that are either local or remote. EPE also has facilities for linking graphical objects to external databases and Web resources, and is capable of reproducing most existing graphical notations and visual representations of pathway maps. In summary, EPE provides a highly flexible tool for combining visualization, editing, and database manipulation of information relating to biological networks. EPE is open-source software, distributed under the Eclipse open-source application platform license.

Introduction

Given the ever-increasing complexity of biological networks, such as those represented by metabolic, signal transduction, and gene transcription regulation pathways, visualization techniques are critically important for further development in integrative systems biology and our understanding of cellular behavior [1–6]. While significant progress has been made in this area, a number of limitations still exist in current software pathway editors.

Biological networks are used in several domains of modern biology and medicine. Several of these domains have a long history, and researchers in these areas have their own established conventions with respect to data visualization. Additionally, the importance of visual representations alone is sometimes overestimated. Attempts to visualize all information in a single big picture often result in a confusing diagram that is difficult to

interpret. Moreover, at the pictorial level, it is not possible to show all available information in a single diagram.

While large overview maps of biological interactions are useful for identifying missing information and understanding global systems behavior, it is very difficult to develop, annotate, and verify data using these maps. Hierarchical organization of information and active (hyperlinked) content can address these difficulties.

A common problem with biological networks editors occurs because an application is often tightly linked to the proprietary database or to the notation developed by its authors. This makes it difficult to extend the information stored within a model and integrate it with information from other proprietary software.

In the Edinburgh Pathway Editor (EPE), we have tried to address all of these important concerns in creating a tool that can be used by biologists and can also provide the required level of mathematical abstraction

©Copyright 2006 by International Business Machines Corporation. Copying in printed form for private use is permitted without payment of royalty provided that (1) each reproduction is done without alteration and (2) the *Journal* reference and IBM copyright notice are included on the first page. The title and abstract, but no other portions, of this paper may be copied or distributed royalty free without further permission by computer-based and other information-service systems. Permission to *republish* any other portion of this paper must be obtained from the Editor.

0018-8646/06/\$5.00 © 2006 IBM

Table 1 Pathway editors.

<i>Criteria</i>	<i>EPE</i>	<i>CellDesigner</i>	<i>TERANODE</i>	<i>Bio Sketch Pad</i>	<i>JDesigner</i>
Last modification date	02.2006	9.2005	11.2005	03.2004	03.2006
Operating system	Any	Any	Win32/Linux**	Any	Win32/Linux
Language	Java**	Java	Java	Java	Delphi
Availability	Free	Free	Commercial	Free	Free
Metabolism	+	+	+	+	+
Signaling	+	+	+	+	+
Gene regulation	+	+	+	+	+
Annotation	+	+ (limited)	+	–	+ (limited)
Change annotation properties	+	–	+	–	–
Change visual presentation	+	+	+	–	+
Import	–	SBML	SBML2.1	SBML2.1	SBML
Export	SBML	SBML	SBML2.1	SBML2.1	Many
Externally controlled vocabulary	+	±	±	–	–
Hierarchy	+	–	+	–	–

and sophistication to create mathematical models of biological networks.

Note that the field of biology is developing rapidly, and new intracellular biomolecules, such as microRNAs that control gene expression, have recently been discovered. These emerging entities are providing a paradigm shift in our understanding of biological systems, and EPE is ready to accommodate new types of networks associated with such discoveries.

Existing pathway editors

We compared EPE with a number of existing pathway editors. **Tables 1, 2, and 3** depict our best knowledge of tool features at the time this paper was written; tool makers frequently offer new features with successive versions and updates. The plus and minus symbols in the tables indicate that a particular pathway editor has or does not have a particular feature. The following criteria were selected to compare the editors from the point of view of a heuristic model builder:

- *Last modification date* – date of the last software release. For example, the oldest modification date is for NetBuilder, which was released in 2003. NetBuilder is an interactive graphical tool for representing and simulating genetic regulatory networks in multicellular organisms. All other applications are updated quite often.
- *OS* – list of operating systems supported. Most of the software is based on Java and is essentially operating-system-independent.

- *Language* – the programming language in which the software was constructed.
- *Metabolism/signaling/gene regulation* – whether the application is able to visualize and handle models of a biological network from distinct domains.
- *Annotation* – the ability to annotate network objects.
- *Change annotation properties* – the ability to edit the annotation in accordance with a procedure created by the user. Examples include the ability to change the ontology (which may be thought of as an agreed-upon categorization or naming convention) or to add a new property for storing specific experimental data.
- *Change visual presentation* – the ability to modify visual properties of the diagram—for example, the shape, color, and size of an object.
- *Externally controlled vocabulary* – the ability to reference external data sources, using a dictionary of concepts such as subcellular locations, cell types, and compounds, in order to facilitate data dissemination and appropriate mapping of names to concepts. Vocabularies are often curated by a group of experts in related domains. A plus/minus symbol in Table 1 indicates software that has the ability to reference externally controlled vocabulary, but using a predefined list of references.
- *Hierarchy* – the ability to divide a model into a hierarchically organized set of sub-models.

Note that a minus sign in the Import column indicates that an application cannot import an external format

Table 2 Pathway editors, continued from Table 1.

<i>Criteria</i>	<i>BioUML</i>	<i>BioTapestry</i>	<i>Pathway Builder 2.0</i>	<i>NetBuilder</i>	<i>VitaPad</i>
Last modification date	5.2005	7.2005	11.2005	6.2003	3.2005
Operating system	Any	Any	Any	Win32	Any
Language	Java	Java	Flash	C++	Java
Availability	Free	Free	Commercial	Free	Free
Metabolism	+	—	—	—	+
Signaling	+	—	+	+	—
Gene regulation	+	+	—	+	—
Annotation	—	—	By name	—	+
Change annotation properties	—	—	No model	—	—
Change visual presentation	—	—	No model	—	—
Import	SBML	CSV	—	—	TXT
Export	SBML	SBML	Image	—	Image
Externally controlled vocabulary	—	—	—	—	KEGG
Hierarchy	—	+	—	—	—

Table 3 Pathway editors, continued from Tables 1 and 2.

<i>Criteria</i>	<i>PathwayLab</i>	<i>PATIKA</i>	<i>PathwayStudio</i>
Last modification date	11.2005	10.2005	11.2005
Operating system	Win32	Any	Win32
Language	C++, requires Visio	Java	C++
Availability	Commercial	Free	Commercial
Metabolism	+	+	+
Signaling	+	+	+
Gene regulation	+	+	+
Annotation	+	+	+
Change annotation properties	—	—	+
Change visual presentation	+	—	+
Import	—	—	—
Export	SBML2.1	SBML	—
Externally controlled vocabulary	—	+	+
Hierarchy	—	+	—

such as the Systems Biology Markup Language (SBML), which is a computer-readable format for representing models of biochemical reaction networks. For the purpose of pathway-editor comparison, we have selected applications that are mostly SBML-compatible, at least with respect to Export features, because SBML is a *de facto* standard in the field of biological modeling. Twelve applications have been included in the comparison: CellDesigner** [7], TERANODE** [8],

Bio Sketch Pad [9], Systems Biology Workbench (SBW) JDesigner [10], BioUML [11], BioTapestry [12], Pathway Builder [13], NetBuilder [14], PathwayLab [15], VitaPad [16], PATIKA [6], PathwayStudio** [17], and EPE. The phrase “no model” in Table 2 refers to the fact that a tool may provide a means for visualization, but the tool does not offer a convenient means to share this information through a model. It turns out that ten of the reviewed editors are model-specific, and eight of

them provide a simulation engine. Of these, BioTapestry, BioUML, and VitaPad have limitations in the description of model elements. For example, VitaPad is not adaptable for signal transduction networks, and BioTapestry restricts itself to use with gene networks alone.

Moreover, all of the applications, with the exception of PATIKA, are unable to work with hierarchical data. Consequently, the process of model generation for the whole organism, even for one as small as the mycoplasma *M. genitalium* with 600 genes, becomes very complex and error-prone. Finally, these editors provide the user with only one text field for the purpose of annotating the displayed objects. These comparisons highlight the need for extending annotation functionality, which would be especially useful for working with multiple literature sources to check the validity of a model. For most of the reviewed pathway editors, the SBML import functionality is also very limited, although most are able to export SBML models.

Tables 1–3 emphasize bioinformatic aspects of several pathway editors, but the reader should note that we have not included every relevant tool in our tables and that some of the tools mentioned herein were designed to emphasize other systems biology applications and not necessarily bioinformatics. However, the general lack of versatility and extendibility is often a shortcoming of many available applications.

Feature summary of EPE

We have sought to overcome these aforementioned limitations by developing EPE as an Eclipse standalone application. Eclipse (www.eclipse.org) is an open-source community whose projects provide a vendor-neutral open development platform and application frameworks for building software. An Eclipse standalone application is software that uses a limited part of the Eclipse platform as a basis for the user interface and for the implementation of low-level system processes. EPE uses the Eclipse Graphical Editing Framework (GEF) as the basis for its drawing functionality. GEF allows developers to create a rich graphical editor from an existing application model.

Generally speaking, Eclipse allows developers to maintain a convenient balance between software extensibility and maintainability through the use of extension points that allow researchers to develop specialized plug-in software for scientific computing. One metaphor for describing extension points and extensions is a headphone jack and headphone wire. The headphone jack is the extension point. The headphone with its wire that plugs into the jack is the extension. All types of headphones with their wires can plug into the headphone jack if the jack and wire are built to fit together properly. When a software application must accommodate

other plug-ins in order to extend its functionality, the application will declare an extension point. The extension point declares a contract (for example, a Java interface) to which extensions must conform. This allows software plug-ins built by different research groups to interact in an easy fashion. Further information on Eclipse extensions and extension points is available from the book *Official Eclipse 3.0 FAQs* [18].

EPE uses a small number of *basic objects*, that is, discrete items that can usually be represented in a visual fashion and that have a number of properties such as a name for each object. These objects illustrate the main concepts of a biological network. Let us now consider the primary, high-level concepts of EPE, which are capitalized and italicized in the following sentences. *Shapes* are biological items or subsystems, which are generally treated as “black boxes” with a number of *Ports* for interfaces. (The term *black box* implies that a researcher does not need to know the details of the system visualized by the Shape object in order to make use of them at this level of abstraction.) *Processes* are concerned with the visualization of a sequence of events, for example events in a biochemical reaction or protein interaction. *Links* are used to represent any pairwise relation between elements, such as Shapes, that are represented in a diagram. Links include “identity” or “act on” relationships, when two objects represent the same entity or when one object regulates or modifies the activity of the other, respectively. *Labels* are items that represent textual information and incorporate links to other maps and resources. The EPE concept of *Context* separates metadata and standards for visualization from pathway maps and pathway data, and it allows one to tune the “drawing palette” for the selected type of map. A drawing palette includes a set of drawing objects and tools that are reminiscent of the erasers, paintbrushes, or pencils that are often available in standard drawing packages. Given the notion of context, a researcher can conveniently create a new object with special customized properties. The concept of Context refers to a collection of objects, their properties, and their default values. It helps users create new objects on the basis of existing ones. The context property editor provides great flexibility in how information is stored and visualized.

Other types of information are not shown on pictorial pathway maps. For instance, EPE captures the provenance of relationships that include literature annotations and links to databases that corroborate the relationships depicted on the map. This information is normally stored in a database or as annotation comments that are supplementary to the map. EPE allows users to customize the list of object properties and to store these data within the object. The data are visualized by means of the graphical representation of linked pages or by pop-

up windows. The general concept of properties allows the system to refer to values of other properties that are already defined, or even to properties of other objects, with the help of the simple language and conventions that we provide, and this greatly reduces the effort needed to update the information and to maintain its consistency. All objects can be marked as searchable, and it is possible to search for the information stored in the properties of those objects. Researchers can find objects on that map that have specific properties; they can also find objects that are associated with specific data. EPE provides two types of search facilities. The first is a simple mode that uses substring search of all objects in the map, folder (containing a collection of maps), or specified data source (all maps in a database). An advanced mode allows users to restrict searching according to the specific property of the specific object type. A typical graphical user interface for search is shown in **Figure 1**. As an example, a researcher might use this search facility to obtain a list of all pathways in a database that involve the coenzyme NADPH, or to graphically highlight all coenzymes in a single particular pathway. Moreover, with the advanced search function, it is possible to find all pathways that involve NADPH, open other maps that contain diagrams for these pathways, and then visually locate all NADPH species. It is also possible to change the visual characteristics of these objects in order to highlight them.

With simple search it is possible to locate all objects for which the searchable properties contain a specified string, so not only will NADP be located, but also NADPH and all enzymes containing NADP in the description of their function.

The information for model annotation comes from external databases, such as SwissProt** or GenBank**, scientific papers, which are generally referenced by a PubMed** ID; and an annotator's or researcher's own knowledge and expertise. One common referencing example or parameter is the PubMed PMID number [19]. PubMed is the National Library of Medicine search service, which provides access to more than 16 million citations. PMID is an acronym for PubMed Identifier, which is a unique number that is assigned to each PubMed citation of life sciences and biomedical scientific journal articles. Other referencing examples or parameters include GenBank accession numbers [20], Protein Data Bank (PDB) record IDs [21], Swiss-Prot IDs [5], and Enzyme EC numbers [22]. EPE provides a default implementation of a link to external databases and publishes corresponding Eclipse extension points. EPE allows data providers to enhance default behavior with vendor-specific features such as complex search facilities or automatic field filling.

EPE permits users to create hyperlinks between pathway maps and provides a means for organizing

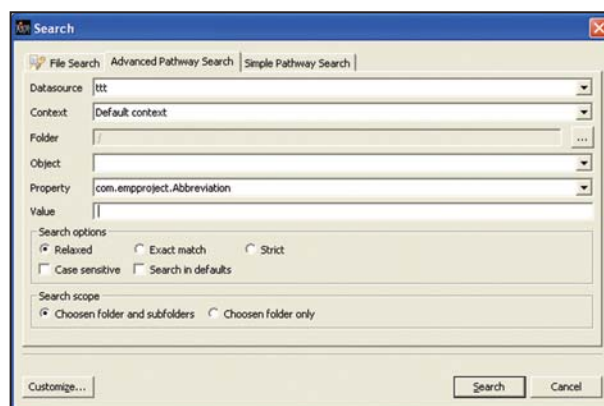


Figure 1

Advanced search window.

information as a hierarchy of maps. This simplifies the process of analysis and verification by allowing researchers to focus on a small subset of data that may exist in a very large model.

EPE supports the addition of reaction kinetic information that is associated with biological processes. Additionally, EPE stores information about maps in a relational format. Databases such as Apache** Derby [23], MySQL** [24], and Oracle** are now supported for internal persistent storage. Apache Derby is treated as local storage, and MySQL and Oracle may be used for enterprise sharing. Any other type of persistent storage may be implemented as a plug-in through the published extension points.

EPE supports data sharing and distribution through Java Database Connectivity (JDBC**) and also through an open Extensible Markup Language (XML**) export format. The XML files can be used for archiving and backup. Pathway diagrams that are created by researchers can be saved as a model in SBML format, or exported to common image formats including JPEG**, PNG, WMF, and SVG**. They may also be converted to a fully functional and hyperlinked HTML tree that is useful for sharing the diagrams via the World Wide Web and providing viewing capabilities for the users who may not have the full editing EPE software installed. If necessary, it is also possible to generate the full list of reactions present on the diagram, in simple text form, or, conversely, to automatically import a reaction from an ASCII text string to the map.

A Java-based architecture powered by Eclipse makes it possible to run EPE on different platforms, from Mac to UNIX** workstations. XML-based export and support of Oracle-based RDBMS (relational database management system) storage systems allow team development of large-scale models.

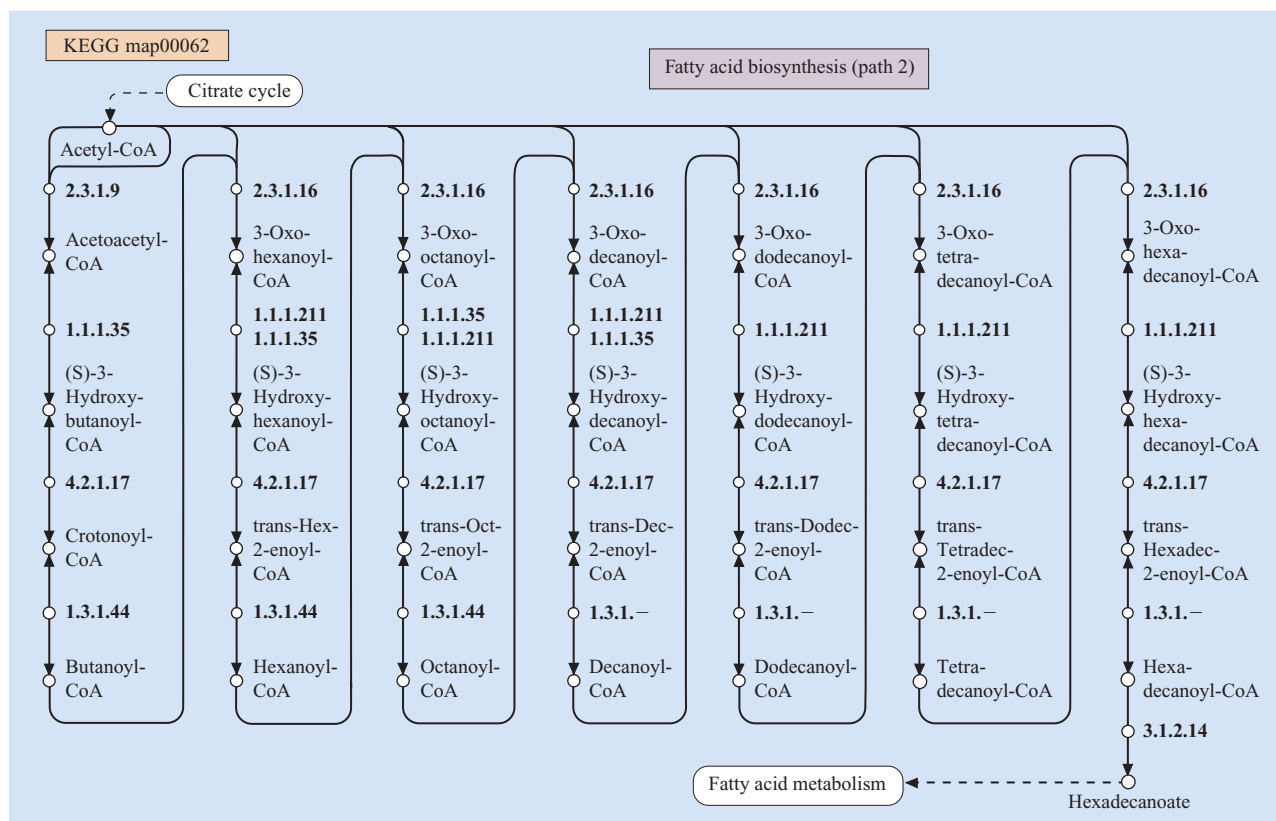


Figure 2

Pathway in KEGG visual style.

A flexible visual representation

One of the main advantages of EPE is the ability to represent information using virtually any visual notation. The key object that facilitates this ability is the *Context* of the map. The Context can serve as a main repository for storing map metadata. It defines a list of objects that are used for drawing a map, the visual and data properties of these objects, links to the external databases, and default values. Later in this paper, we compare several contexts, and maps created using different contexts.

The Kyoto Encyclopedia of Genes and Genomes (KEGG) [25] provides well-known visualizations of the metabolic pathways in its collection of pathways. KEGG is a relatively simple format comprising three children of the Shape object and two types of Labels. The Compound object extends the concept of Shape and visually represents chemical compounds with small circles. It stores the name of the molecule and what we call a KEGGCID (KEGG Compound ID) property to reference the KEGG molecule ID. The special Label is used to show the name of the compound and allows the system to open KEGG compound descriptions on the

basis of the KEGGCID property. Biochemical reactions are visualized by an ordinary Process object with additional properties to store EC (Enzyme Commission) numbers used to reference KEGG process descriptions. A special type of Shape is used to indicate literature references, quality control, and level of confidence of the pathway in the defined organism or tissue. In addition, literature annotation can be stored in the properties.

Only the main compounds are shown on a KEGG map; see for example the map in Figure 2, which represents the biochemical process of fatty acid biosynthesis. This map was created in EPE and then exported to WMF (Windows Metafile Format) and converted to EPS (encapsulated postscript). Note that no H₂O, ATP, or NADPH molecules are shown, even though such molecules are involved in those enzyme processes. This lack of detail sometimes misleads scientists, especially when they are analyzing cellular data without paying much attention to the biology of the specific domain. A profound analysis of the consequences of such a shallow approach is presented in [26].

Another format used to represent metabolic pathways is produced by the EMP-MPW-WIT [27, 28] cluster of

systems and databases developed by researchers at the Argonne National Laboratory and the Russian Academy of Sciences, in collaboration with researchers from other organizations (**Figure 3**). WIT is an integrated system for the support of genetic sequences, comparative analysis of sequenced genomes, and metabolic reconstructions from the sequence data. EMP is the largest publicly available database of enzymes and metabolic pathways in the world. MPW is a collection of metabolic and functional diagrams. EMP pathway diagrams contain all of the compounds involved in biochemical reactions: cofactors, energy and charge supply molecules, water, and gas molecules. In this case, biochemical processes are visualized as a set of Process objects with additional EC and EN (enzyme reaction name) properties to identify the catalyzer (protein) of the process. A special kind of Label has been introduced to show this information on the map. It links the Process object to the IUPAC Enzyme Nomenclature [29] page via the EC number.

All of the compounds on the EMP map are connected to Ports of the Process objects. Ports that represent the same compound are linked by a Link of the “equal to” type. This Link type is able to transfer values of the properties between the connected objects, and it provides “equality requirements” for objects on both sides of the link that may share certain properties such as compound name or ID to represent the same chemical compound involved in a number of reactions. This feature is analogous to the so-called “node clones” in some of the other editors, and can be very useful in reducing human error and increasing the drawing rate of large maps with high connectivity.

In Figure 3, which shows the same pathway as Figure 2, note that all compounds on the EMP map are color-coded. The small green and red circles respectively indicate substrates and products. The type of compound represented is defined by the color of its label. Backbone molecules of the pathway (molecules involved in carbon or/and nitrogen transfer) are colored cyan. Cofactors and energy-supply molecules are yellow. Molecules that exist in abundance, such as H₂O and CO₂, are referred to as solution molecules and colored white. All compound labels can be linked to the dictionary of small (less than 2000-Dalton) biochemical compounds such as those in the KEGG Ligand database. Some objects have PubMed links to provide information or evidence for a pathway in a specific subcellular location in specific cell types of defined organisms. Figure 3, like other pathway diagrams in this paper, is rendered using EPE.

As mentioned earlier, EPE has the ability to organize knowledge in a hierarchical way as outline maps (**Figure 4**). The colored circles represent ports and represent inputs and outputs of the subsystems. Cellular

subsystems are denoted as black boxes with defined inputs and outputs. The detailed map for a pathway is available by clicking on the text of the name in a box.

EPE can be used to visualize signal transduction and genetic regulation maps. **Figure 5** shows one of the simplest map representations that researchers can use to represent protein interaction events. Colors are used to help researchers identify subcellular locations. Ellipses on this map represent proteins in the phosphatidylinositol-3,4,5-trisphosphate pathway. Abbreviated protein names are shown inside ellipses. Hyperlinks provide links to databases such as the Swiss-Prot protein database. All protein–protein interactions, modifications, and other events in this notation are visualized by the previously mentioned Process object. In this case, Process may have an additional type of object linked to it (a regulator) that does not exist in metabolic maps. A regulator object links with a Process by an arrow pointing to the Process. The properties of this object include type of regulation, literature details, and experimental proof for the regulation event.

It is possible to create artistic forms of the diagrams which are similar to those of Pathway Builder [12] or BioCarta** [30]. The EPE pathways contain full-featured models, including mathematical formulas which do not appear in the picture (**Figure 6**). The artistic pathway diagram for the epidermal growth factor (EGF) pathway can easily be created from a standard pathway diagram by replacing shapes with graphical images or icons. The mathematical model behind the diagram remains unchanged after the conversion.

Another biological domain covered by EPE is genetic regulation or gene networks (**Figure 7**). A gene network is a collection of DNA segments in a cell that interact with each other and with other biomolecules to govern the rates at which genes in the network are transcribed into mRNA. The bottom line of shapes in this figure provides information on genetic information for the fatty acid synthase gene. The top line of rectangles represents domains and important amino acids in the protein that is a product of this gene. Several new Shape children have been introduced to the Context to visually represent the distribution of genes on the chromosome map. The Promoter objects, labeled PI and PII, represent the promoter locations on the chromosome. Intron and Exon objects respectively store information on introns (DNA regions that do not code for proteins) and exons (DNA coding regions). Exons are represented as rectangles on the bottom line of the figure, and the || symbol represents introns. Intron and exon objects store lists of transcriptional factors and alternative splicing sites in addition to the name, the chromosome location, and references to nucleotide databases and literature.

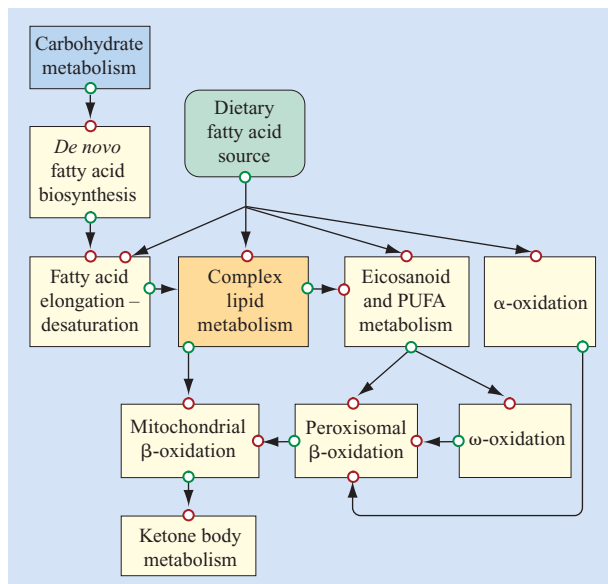


Figure 4

WIT outline diagram.

Clearly, we have not given a complete list of all properties needed to describe a genetic regulatory network. However, as mentioned earlier, new types of data and properties can easily be added to the corresponding objects in the context.

As a final visualization, **Figure 8** shows a part of the complex compartmental genetic network for the interferon pathway. The very small squares represent input and output ports. The term *compartment* refers to different regions in the cell, such as the cytosol, endosome lumen, and endosome membrane. In this representation, a great number of different types of interactions between objects have been introduced. The context of this map (**Table 4**) contains multiple objects and properties. During the development of this model, a special graphical notation was created that contained thirty objects and was implemented in the new Context.

Genes, Proteins, and Compartments are children of the Shape object in the sense that these objects were created by inheritance from Shape. The Compartment object is used to represent different parts of the cell. It is very important for the model to clearly define the cellular location of other objects on the map. The protein translocation from one compartment to another is usually accompanied by a protein modification event. Each object on the interferon map has a Location property which in turn refers to a Compartment object.

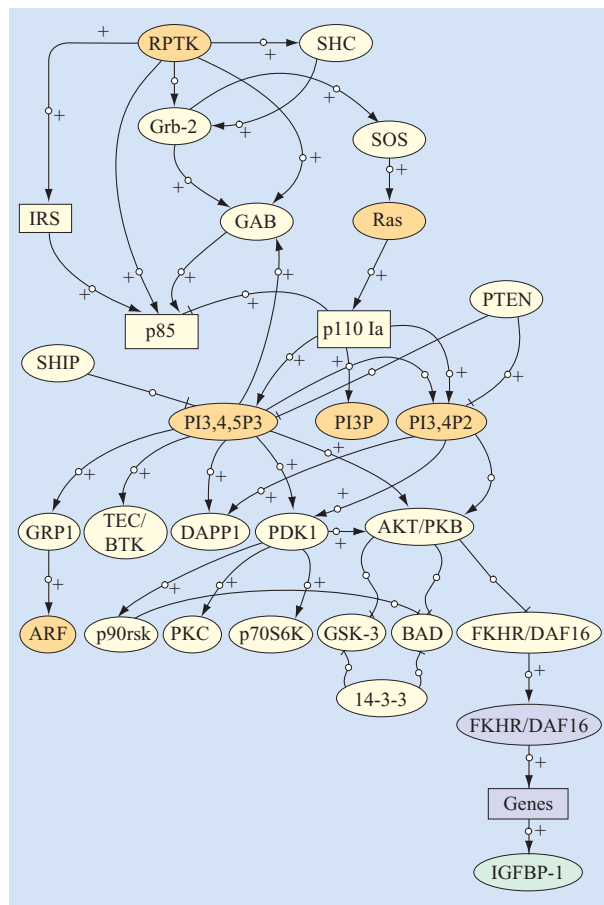


Figure 5

Signal transduction pathway.

In EPE, Protein and Gene objects contain references to corresponding databases and literature annotation. All objects that reference the same paper can easily be found by search procedures that are provided. This feature allows researchers to simplify the verification of the model and the literature curation process. Pairwise relationships, such as those associated with protein translocation or process inhibition, are listed as children of a Link object.

In addition, the Protein Interaction and Protein Modification objects are also represented as children of the Process object. The Protein Interaction function is denoted by a small black circle with two kinds of links. Ordinary Links point to the Substrate and the Product objects that are involved in the interaction. Regulatory Links point away from an object, such as an activator or inhibitor, that could change the rate of a process. An Activation Link ends with an arrowhead, and

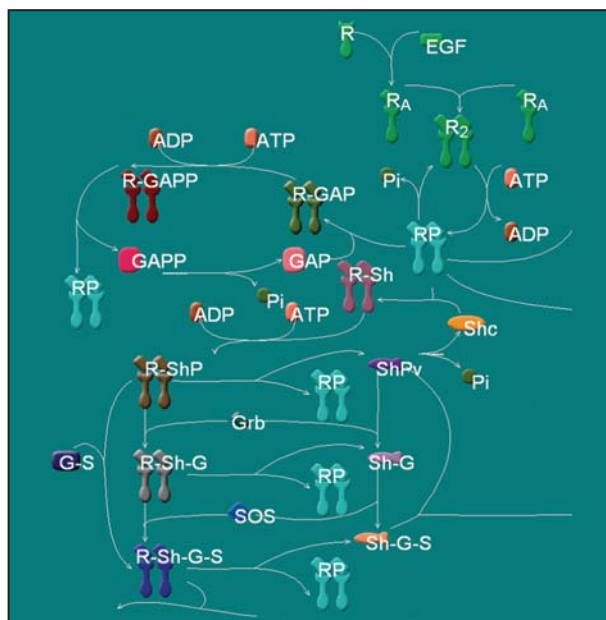


Figure 6

Portion of signal transduction pathway in artistic form.

an Inhibition Link ends with a bar head.¹ Different modification processes are represented as open circles labeled with the letters P for phosphorylation, M for methylation, and A for acetylation.¹ Very small red and green squares indicate input and output ports.

EPE provides a special type of object known as a Gate object to represent the inheritance hierarchy. The Gate object is a child of the Process object. It is a high-level logical object, so it cannot be added to the map by itself. The Gate object is a parent for other types of Gates, which are the And Gate, Or Gate, and Not Gate. These gates help to reduce the complexity of the map through the addition of logical functions to regulation links. For example, instead of drawing two inhibition links to the protein binding process from two different proteins, it is easier to add an inhibition link from an Or Gate to the protein binding process, and then to add links from all known inhibitors to this Or Gate. Thus, the Gate object is an example of the kind of a complex organization of information that is supported by EPE.

Conclusion

EPE belongs to a new generation of software tools that will provide the scientific community with the necessary computational aid to manage biological data and perform comparative analyses between such data. The current progress in “omics” technologies is leading to vast

Table 4 List of objects in the Interferon map context, a portion of which is shown in Figure 8. This list indicates all of the objects used to create the Interferon diagram.

<i>Object of the map</i>	<i>Base editor object</i>
Protein	Shape
Gene	Shape
mRNA	Shape
Output	Port
Input	Port
Protein interaction	Process
Phosphorylation	Process
Acetylation	Process
Methylation	Process
Ubiquitination	Process
Mirror	Process
And Gate	Process
Or Gate	Process
Addition	Process
Synergy	Process
Interaction	Link
Stimulation	Link
Inhibition	Link
Total inhibition	Link
Translocation	Link
Degradation	Link
Cleavage	Link
Reverse action	Link

quantities of information that researchers will need to analyze in order to obtain a better understanding of biological phenomena. New data are also being generated at different biological levels—for example, at the phenotype, intramolecular, and intercellular levels. Pathway maps provide a means to organize multidimensional views of a wealth of information and a means to assemble known and novel network characteristics. Pathway generation and analysis has the potential to lead to novel pathway biomarkers, predict possible drug adverse effects, and ultimately reduce the time required for drug development.

With these goals in mind, we have developed the computer-assisted design tool described in this paper. The tool is highly flexible and supports an extendable application for manual generation, annotation, and visualization of biological networks of unlimited size and from different biological domains. The metadata-based

¹These elements do not appear in the part of the interferon pathway shown in the figure.

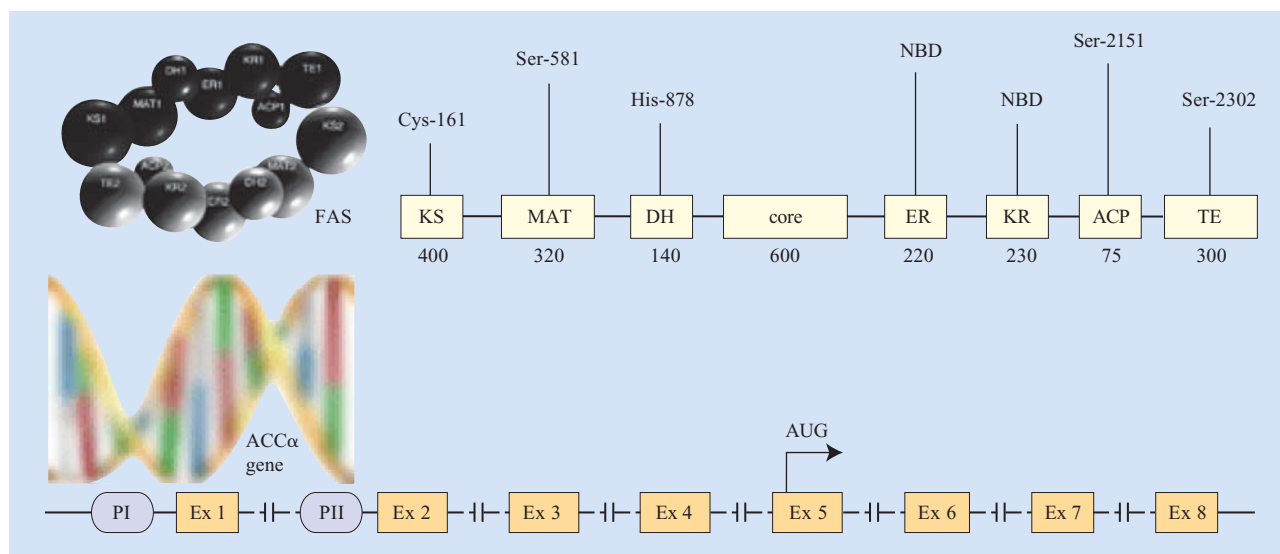


Figure 7

Genetic network of rat fatty acid synthase.

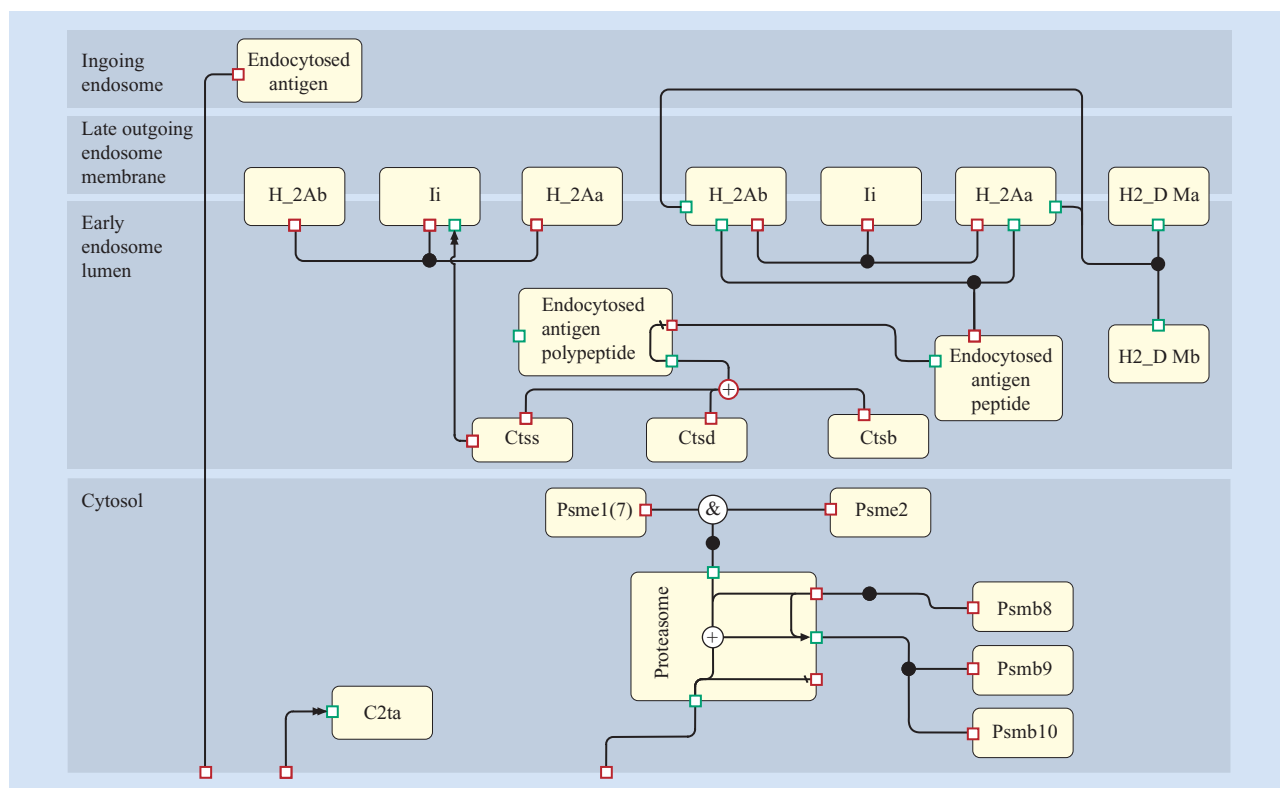


Figure 8

Part of interferon pathway diagram.

architecture makes it easy for users to develop their own data structures to support data-validation and knowledge-extraction procedures. The flexible visualization strategy allows researchers to tune the visual representation of a map for a variety of desired graphical notations. Object-linking techniques support the hierarchical presentation of information. Most significantly, EPE makes pathway maps and models more understandable and maintainable. We hope that EPE will help to address the growing need for systems biology tools suitable for biologists [31]. EPE is a first step toward the new generation of systems biology software that will provide a seamless, transparent front-end interface for theoreticians and experimentalists alike. The software, tutorials, and demonstrations are available at the EPE Web site <http://www.bioinformatics.ed.ac.uk/epe/>.

Acknowledgments

We thank Eugene Selkov and Eugene Selkov, Jr., as well as GlaxoSmithKline, Hugh Spence, and Frank Tobin in particular for ideas, support, useful comments and feedback during the initial stage of EPE development.

**Trademark, service mark, or registered trademark of Sun Microsystems, Inc., Linus Torvalds, The Systems Biology Institute, Teranode Corporation, Ariadne Genomics, the National Library of Medicine, National Institutes of Health, Swiss Institute of Bioinformatics, Apache Software Foundation, MySQL AB, Oracle Corporation, the World Wide Web Consortium, Independent Joint Photographic Experts Group, The Open Group, and BioCarta, Inc. in the United States, other countries, or both.

References

1. D. Cook, J. Farley, and S. Tapscott, "A Basis for a Visual Language for Describing, Archiving and Analyzing Functional Models of Complex Biological Systems," *Genome Biol.* **2**, No. 4, 12.1–12.10 (2001).
2. H. Kitano, A. Funahashi, Y. Matsuoka, and K. Oda, "Using Process Diagrams for the Graphical Representation of Biological Networks," *Nature Biotechnol.* **23**, No. 8, 961–966 (2005).
3. K. Kohn, "Molecular Interaction Maps as Information Organizers and Simulation Guides," *Chaos* **11**, No. 1, 84–97 (2001).
4. K. Kohn, "Molecular Interaction Map of the Mammalian Cell Cycle Control and DNA Repair Systems," *Mol. Biol. Cell* **10**, No. 8, 2703–2734 (1999).
5. C. Wu, H. Huang, L. Arminski, J. Castro-Alvares, Y. Chen, Z. Hu, R. Ledley, K. Lewis, H. Mewes, B. Orcutt, B. Suzek, A. Tsugita, C. Vinayaka, L. Yeh, J. Zhang, and W. Barker, "The Protein Information Resource: An Integrated Public Resource of Functional Annotation of Proteins," *Nucl. Acids Res.* **30**, No. 1, 35–37 (2002).
6. E. Demir, O. Babur, U. Dogrusoz, A. Gursoy, G. Nisanci, R. Cetin-Atalay, and M. Ozturk, "PATIKA: An Integrated Visual Environment for Collaborative Construction and Analysis of Cellular Pathways," *Bioinformatics* **18**, No. 7, 996–1003 (2002).
7. The Systems Biology Institute, "CellDesigner: A Modeling Tool of Biochemical Networks"; see <http://celldesigner.org/>.
8. Teranode Corporation, "Design Automation for Life Sciences"; see <http://www.teranode.com/>.
9. Biocomputation Group, University of Pennsylvania, "Bio Sketch Pad"; see http://www.cis.upenn.edu/biocomp/new_html/biosketch.php3.
10. Keck Graduate Institute, "Keck Computational Systems Biology"; see <http://www.sys-bio.org/>.
11. Biosoft, "BioUML Framework for Systems Biology"; see <http://www.biouml.org/>.
12. Institute for Systems Biology, "BioTapestry"; see <http://labs.systemsbiology.net/bolouri/software/BioTapestry/>.
13. Protein Lounge, "Pathway Builder Tool"; see <http://www.proteinlounge.com/pathwaybuilder.asp>.
14. NetBuilder Development Team, "NetBuilder Home"; see <http://strc.herts.ac.uk/bio/maria/NetBuilder/>.
15. InNetics, "PathwayLab"; see http://innetics.com/pathwaylab_overview.htm.
16. M. Holford, N. Li, P. Nadkarni, and H. Zhao, "VitaPad: Visualization Tools for the Analysis of Pathway Data," *Bioinformatics* **21**, No. 8, 1596–1602 (2005).
17. Ariadne Genomics, "PathwayStudio"; see <http://www.ariadnegenomics.com/products/pscentral/>.
18. J. Arthorne and C. Laffra, *Official Eclipse 3.0 FAQs*, Addison-Wesley Publishing Co., New York, 2004.
19. U.S. National Institutes of Health (NIH), "PubMed Central"; see <http://www.pubmedcentral.nih.gov>.
20. D. Benson, I. Karsch-Mizrachi, D. Lipman, J. Ostell, B. Rapp, and D. Wheeler, "GenBank," *Nucl. Acids Res.* **30**, 17–20 (2002).
21. Research Collaboratory for Structural Bioinformatics, "RCSB Protein Data Bank"; see <http://www.rcsb.org>.
22. Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (NC-IUBMB), "Enzyme Nomenclature"; see <http://www.chem.qmul.ac.uk/iubmb/enzyme/>.
23. Apache Software Foundation, "Apache Derby"; see <http://db.apache.org/derby/>.
24. MySQL AB, "MySQL: The World's Most Popular Open Source Database"; see <http://www.mysql.com/>.
25. H. Ogata, S. Goto, K. Sato, W. Fujibuchi, H. Bono, and M. Kanehisa, "KEGG: Kyoto Encyclopedia of Genes and Genomes," *Nucl. Acids Res.* **27**, No. 1, 29–34 (1999).
26. H. Ma and A. Zeng, "Reconstruction of Metabolic Networks from Genome Data and Analysis of Their Global Structure for Various Organisms," *Bioinformatics* **19**, No. 2, 270–277 (2003).
27. R. Overbeek, N. Larsen, G. Pusch, M. D'Souza, E. Selkov, Jr., N. Kyrpides, M. Fonstein, N. Maltsev, and E. Selkov, "WIT: Integrated System for High-Throughput Genome Sequence Analysis and Metabolic Reconstruction," *Nucl. Acids Res.* **28**, No. 1, 123–125 (2000).
28. E. Selkov, S. Basmanova, T. Gaasterland, I. Goryanin, Y. Gretchen, N. Maltsev, V. Nenashev, R. Overbeek, E. Panyushkina, L. Pronevitch, E. Selkov, Jr., and I. Yunus, "The Metabolic Pathway Collection from EMP: The Enzymes and Metabolic Pathways Database," *Nucl. Acids Res.* **24**, No. 1, 26–29 (1996).
29. International Union of Biochemistry and Molecular Biology, "IUBMB Nomenclature"; see <http://www.chem.qmul.ac.uk/iubmb/>.
30. "BioCarta—Charting Pathways of Life"; see <http://www.biocarta.com>.
31. M. Cassman, "Barriers to Progress in Systems Biology," *Nature* **438**, No. 7071, 1079 (2005).

Received January 13, 2006; accepted for publication February 3, 2006; Internet publication September 19, 2006

Anatoly Sorokin *University of Edinburgh, Appleton Tower, Crichton Street, Edinburgh EH8 9LE, United Kingdom (asorokin@inf.ed.ac.uk)*. Dr. Sorokin is a Research Fellow in the Computational Systems Biology Group at the University of Edinburgh. In 1995 he received an M.Sc. degree from the Moscow Institute of Physics and Technology and in 2000 a Ph.D. degree from the Institute of Experimental and Theoretical Biophysics, Pushchino, Russia. In 1998, Dr. Sorokin began work in the Institute of Cell Biophysics in Pushchino, where he focused on the prediction and analysis of physical properties of DNA.

Kirill Paliy *Gazetnyj Lane, Building 100, Apartment 104, Rostov-na-Donu, Russia (kirill@empproject.com)*. In 1994, Mr. Paliy graduated with an M.Sc. degree in software engineering from the Rostov State University of Railway Transportation. From 2002 to 2006, he worked for EMP Project, Inc. as a software development consultant. He made significant contributions to the current state of the Edinburgh Pathway Editor.

Alexey Selkov *EMP Project, Inc., 1409 Heatherton Drive, Naperville, Illinois 60563 (as@empproject.com)*. Mr. Selko has led the EMP Project effort since 1999. From 1988 to 1998, as a project leader, he built various large networking infrastructures for government and private companies.

O. V. Demin *A. N. Belozersky Institute of Physico-Chemical Biology, Leninskie Gory, Moscow State University, Moscow 119992, Russia (demin@genebee.msu.su)*. Dr. Demin received his M.S. degree and his Ph.D. degree, both in biophysics, in 1992 and 1995 from Moscow State University. In 1996, he began work in the Bioenergetics Department of the A. N. Belozersky Institute of Physico-Chemical Biology. Since 1999, he has been the head of the Kinetic Modeling of Complex Biochemical Systems Group. Dr. Demin's current interests include development of software for the mathematical simulation and analysis of cellular metabolism and regulation.

Serge Dronov *GlaxoSmithKline, Gunnels Wood Road, Stevenage SG1 2NY, United Kingdom (serge.dronov@gsk.com)*. In 1994, Mr. Dronov received his M.S. degree in applied mathematics and computer science from the Moscow Institute of Physics and Technology, Russia; he is currently working toward a Ph.D. degree in biophysics. In 2000, following his postgraduate research at the Institute of Biophysics and work on information systems for the Russian Academy of Science, he joined the Scientific Computing and Mathematical Modeling Group at GlaxoSmithKline's Research and Development organization in the UK, where he worked on large-scale cell simulations. Mr. Dronov's research interests include systems biology, modeling of biological pathways, numerical methods for solving differential equations, optimization techniques, and graph theory algorithms. He has a strong information-technology background and experience in scientific software and database development.

Peter Ghazal *Scottish Centre for Genomic Technology and Informatics, University of Edinburgh Medical School, The Chancellor's Building, Little France Crescent, Edinburgh EH16 4SB, United Kingdom (P. Ghazal@ed.ac.uk)*. Dr. Ghazal is Professor of Molecular Genetics and Biomedicine at the University of Edinburgh Medical School, Adjunct Professor at the Scripps Research Institute, La Jolla, California, and the founding director of the Scottish Centre for Genomic Technology and Informatics.

Before holding these positions, Dr. Ghazal was Associate Professor at the Scripps Research Institute in California and a Fellow of the National Institutes of Health (NIH) in Bethesda, Maryland. He was also a Scholar of the American Leukemia Society. He has several patents in the area of engineering and biology, is a cofounder of two companies, and has published more than 100 refereed papers. Dr. Ghazal's research interests focus on regulatory networks and host pathogen interactions. He is also a member of a number of advisory boards and committees for several industry, academic, and government bodies in the UK, EU, USA, and Canada.

Igor Goryanin *University of Edinburgh, Appleton Tower, Crichton Street, Edinburgh EH8 9LE, United Kingdom (goryanin@inf.ed.ac.uk)*. Professor Goryanin is the Henrik Kacser Chair in Computational Systems Biology and Director of the Edinburgh Center for Bioinformatics at the University of Edinburgh. He is well known as an author of DBsolve, one of the first software packages for modeling and simulation of cellular networks. Dr. Goryanin recently worked at GlaxoSmithKline (GSK), focusing on the areas of bioinformatics and informatics. At GSK, he was head of the Cell Simulations and Pathway Modeling Group. Prior to his affiliation with industry, Dr. Goryanin worked at the Institute of Biophysics and the Russian Academy of Science. He was also a visiting scientist at Argonne National Laboratories. He received his Ph.D. degree in 1995 from the Institute of Biophysics in Russia, where he worked on the automation of multienzyme systems modeling, using a databank of enzymes and metabolic pathways. He received an M.Sc. degree in 1985, as an applied mathematician, from the Computer Science Department of the Moscow Engineering Physical Institute (MEPHI), where he developed numerical methods and algorithms for the analysis of stiff ordinary differential equations. More recently, Dr. Goryanin has established several collaborations between industry and academia, including the first International Alliance on Systems Biology of *E. coli* (IECA), where he is a member of the steering committee. He also holds a position as a visiting professor in the Biochemical Engineering Department of University College, London.